The emergence of systematic argument distinctions in artificial sign languages

Journal of Language Evolution, 2021, 77–98
doi: 10.1093/jole/lzab002
Advance Access Publication Date: 18 June 2021
Research article

OXFORD

The emergence of systematic argument distinctions in artificial sign languages

Yasamin Motamedi [1,*], Kenny Smith [1], Marieke Schouwstra[1,2], Jennifer Culbertson[1] and Simon Kirby[1]

[1]Centre for Language Evolution, University of Edinburgh, 3 Charles Street, Edinburgh, EH3 9AD, UK and [2]Faculty of Humanities, University of Amsterdam, Spuistraat 134, Amsterdam, 1012 VB, the Netherlands

*Corresponding author: Yasamin.Motamedi@ed.ac.uk

Abstract

Word order is a key property by which languages indicate the relationship between a predicate and its arguments. However, sign languages use a number of other modality-specific tools in addition to word order such as spatial agreement, which has been likened to verbal agreement in spoken languages, and role shift, where the signer takes on characteristics of propositional agents. In particular, data from emerging sign languages suggest that, though some use of a conventional word order can appear within a few generations, systematic spatial modulation as a grammatical feature takes time to develop. We experimentally examine the emergence of systematic argument marking beyond word order, investigating how artificial gestural systems evolve over generations of participants in the lab. We find that participants converge on different strategies to disambiguate clause arguments, which become more consistent through the use and transmission of gestures; in some cases, this leads to conventionalized iconic spatial contrasts, comparable to those found in natural sign languages. We discuss how our results connect with theoretical issues surrounding the analysis of spatial agreement and role shift in established and newly emerging sign languages, and the possible mechanisms behind its evolution.

Key words: iterated learning; communication; sign language; silent gesture; spatial reference

1. Introduction

A fundamental requirement placed on human language is the need to indicate the relationship between a predicate andits arguments. For example, to successfully communicate a simple transitive event, it is necessary to convey who was the agent of the action, and who the patient. How did the tools employed by languages for indicating such relationships originate? Emerging sign languages provide invaluable evidence of this process because we can observe the gradual evolution of distinct strategies for indicating the role of participants in a proposition (Senghas and Coppola 2001; Padden et al. 2010; Montemurro et al. 2019). In this article, we will use *artificial* sign languages to examine this evolutionary process in more detail in the lab.

1.1 Predicate–argument relations in natural sign languages

Languages across the world mark linguistic relationships between a predicate and its arguments using tools such as word order (Dryer 2013) and verb agreement (Siewierska 2013). Sign languages also convey the relationship between predicate and argument with word/sign order (Sandler and Lillo-Martin 2006; Leeson and Saeed 2012), though modality-specific constraints have been proposed that lead to different tendencies compared to spoken languages (Napoli and Sutton-Spence 2014; Johnston 2019). However, sign order is not the only tool available to signers, and sign languages employ other, modality-specific tools to denote the semantic and syntactic relationships between predicates and arguments.

The most well-documented of these tools is the use of verbal spatial modulation, or verb directionality, which has been likened to verb agreement in spoken languages (Padden 1990; Liddell 2003; Lillo-Martin and Meier 2011; Mathur and Rathmann 2012). Referents (usually animate arguments) are tracked across discourse using locations indexed in the space around the signer (often with points), with verb forms moving between these locations, or between the indexed locations and the signer's body (Liddell 2003). Spatial modulation of this kind usually affects a subset of verbs in the language, namely transfer verbs like ASK, PAY, and GIVE. An example of how this device works in British Sign Language (BSL) is shown in Fig. 1 (adapted from an example given in Morgan, Barrière, and Woll 2006); in this example, the referential locations of the agent (girl) and patient (boy) are introduced first (as IX_A and IX_B, respectively), and the verb ASK is then modulated to move between these two loci to signify who is asking whom (Morgan et al. 2006). The use of spatial modulation to denote arguments in discourse is attested in most sign languages (Sandler and Lillo-Martin 2006; Mathur and Rathmann 2012), and spatial modulation has been

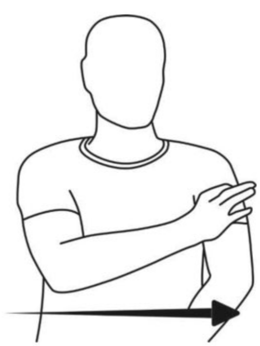


Figure 1. Example of spatial modulation with referential loci. Movement in the image illustrates the movement of the verb sign ASK (bold text) between two referential loci, representing the two third-person arguments.

suggested to be an inevitable consequence of using language in a visual modality (Meier 1990; Aronoff, Meir, and Sandler 2005).

Another, less well-documented use of space to denote predicate–argument relationships is the use of role-shift (also termed constructed action, e.g., Cormier, Smith, and Zwets, 2013). Signers use the orientation and positioning of their own bodies to distinguish between animate arguments in a clause, embodying multiple arguments to differentiate them. The example in (Fig. 2), (adapted from Padden (1990)), illustrates this phenomenon in American Sign Language (ASL), though it has been documented in several other sign languages, such as BSL (Cormier et al. 2013; Cormier, Fenlon, and Schembri 2015) and Nicaraguan Sign Language (NSL), a language <50-years old (Kocab, Pyers, and Senghas 2015). Here, the signer's body first represents the agent of the event (the man swinging his fist) and, following a shift in body orientation, the patient (the man whose cheek is being swung at). In this way, role shift allows a distinction between clause arguments, exploiting the iconicity of the signer's body. While role shift has been likened to the use of quotatives in spoken language, its use in sign languages extends far beyond a quotative function (Pfau and Quer 2010; Lillo-Martin 2012) and comprises a core expressive tool in many sign languages (Quinto-Pozos 2007; Ferrara and Johnston 2014).

Both spatial modulation and role-shift represent modality-specific uses of space to distinguish between clause arguments, using iconic representations based on the affordances of the signer's body to represent grammatical relationships. Indeed, spatial modulation is often used in conjunction with role-shift (Fenlon, Schembri, and Cormier 2018). In both cases, iconic forms must be reanalyzed in terms of grammatical features, a task which appears to be nontrivial. For example, error-free use of spatial modulation and role-shift may not occur until a relatively late age (Lillo-Martin,1999; Morgan, Herman, and Woll 2002; Chen Pichler 2012). Furthermore, there is evidence that the systematic use of contrastive space does not emerge immediately in new sign languages (Meir et al. 2007; Padden et al. 2010; Vos 2012; Montemurro et al. 2019), calling into question the universality of spatial modulation systems. Research into Israeli Sign Language (ISL) and Al-Sayyid Bedouin Sign Language (ABSL), both young sign languages, indicates that neither had a system of spatial modulation in early generations of signers (Meir et al. 2007; Padden et al. 2010; Meir 2012) and that role-shift does not seem to be present in ABSL (Padden et al. 2010). Studies of NSL have demonstrated similar findings. Consistency in the production and perception of spatial modulations emerges by the second cohort (Senghas and Coppola 2001; Senghas 2003) but continues to systematize through the third cohort (Flaherty 2014; Montemurro et al. 2019). Similarly, Montemurro et al. (2019) found a gradual increase in the use of role-shift over cohorts, with signers only using role-shift consistently by the third cohort.

Furthermore, both ABSL and NSL demonstrate gradual change in *how* signers use space. In both languages, there is a change over generations (or cohorts), from preferring spatial modulations enacted perpendicular to the signer's body, to a preference for spatial modulations parallel to the signer's body (Padden et al. 2010; Montemurro et al. 2019). In the first case, occurring on the *z*-axis (see Fig. 3), the signer's own body acts as an argument, usually as the agent, with an indexed location acting as patient/recipient. Later in the development of these languages, the *x*-axis is employed such that the signer's body is no longer central to the modulation; referential loci index arguments and verbal forms move between these loci, similar to the example shown in Fig. 1. Padden et al. (2010) suggest that the evolution of a

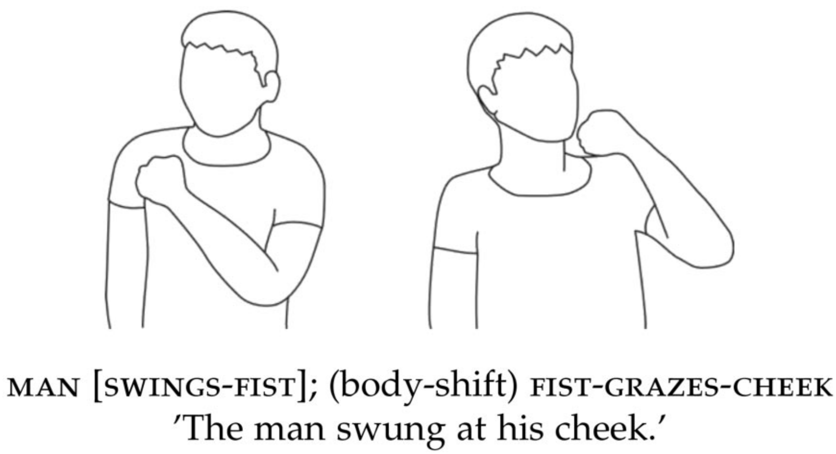


Figure 2. Example of role shift from Padden (1990). The signer first enacts the agent (the man) punching, before re-orienting their body to represent the patient being punched.

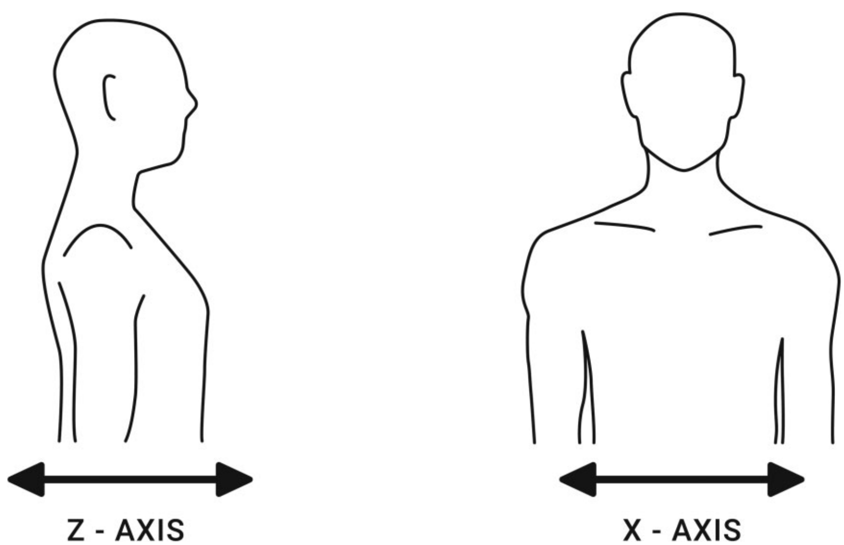


Figure 3. Use of different spatial axes in sign languages. On the left, the *z*-axis moves outwards from the signer's body (here shown in profile). On the right, the *x*-axis sits parallel to the signers body.

spatial system that can contrast two animate participants in different locations removed from the body requires the abstraction of the grammatical concept of *person*. This abstraction is in direct conflict with the potential of the signer's body to first and foremost represent themselves. In other words, the natural first-person form must be co-opted in order to represent nonfirst-person agents (Meir et al. 2013), and thus abstraction from the body allows the signer to avoid this conflict. However, this process of abstraction and grammaticalization may take time to develop in a language.

There are also differences in the semantics of spatial modulation. In older sign languages, such as BSL and ASL, spatial modulation tends to be used with only a subset of signs—namely, verbs of transfer (Padden 1990; Sutton-Spence and Woll 1999; Sandler and Lillo-Martin 2006). Spatial verbs such as MOVE or CARRY can use space to denote the movement of arguments in space but do so by iconically signaling a locative relationship in the real world (i.e., movement of an object between two locations). Verbs of transfer can indicate a direct physical transfer but often represent a grammatical relationship between the predicate and its arguments (e.g., in the sign ASK, the movement between loci represents who asked whom, even if the arguments are not physically in the relative locations as depicted by the referential loci). However, data from ISL and ABSL signers suggest that they do not distinguish between these verb types in earlier generations (Padden et al. 2010). As with the grammatical concept of person, it may be the case that more abstract spatial representations (such that the representation does not actually map on to physical space) may take more time to develop.

Finally, there is a lack of theoretical consensus concerning the grammatical status of spatial modulations in sign languages. One possibility is that spatial modulations are a fully grammaticalized phenomena, parallel to verb agreement in spoken languages, such that a set of abstract features (e.g., person) are triggered when the relevant grammatical conditions hold (Mathur and Rathmann 2012). However, multiple accounts now suggest that spatial grammars show a level of variability that does not concord with this strong agreement-based account. One possibility is that they may represent a noncanonical form of agreement (Lillo-Martin and Meier 2011). Alternatively, they may not represent a case of agreement at all, but rather the combination of morphemic units with gestural points, akin to multimodal constructions of speech and gesture in spoken language (Liddell 2003; Schembri, Cormier, and Fenlon 2018). This latter theoretical claim is supported by data collected from BSL signers (Cormier et al. 2015; Fenlon et al. 2018). These data showed that BSL signers tend to map their spatial modulations onto real-world positioning of animate arguments, rather than using arbitrary positions, and frequently use role-shift to embody animate arguments. The authors interpret their results as showing that spatially modulated constructions do not represent a clearly abstract grammatical use of space.

In summary, while a large number of sign languages do exhibit spatial modulation and role-shift (Sandler and Lillo-Martin 2006), they are not obviously present in at least some newly emerging sign languages, and the extent to which spatial modulation is fully grammaticalized in more established systems is unclear. Here, we undertake the first experimental exploration of how systems that exploit the iconic affordances of signer's bodies and the signing space might emerge, and how these initially gradient gestures might be integrated into a structured linguistic system over time. Having the level of control that an experimental setting offers enables us to generalize over different iconic systems that have a comparable starting point, and analyze how they develop through cultural transmission.

1.2 The evolution of complex linguistic structure in the lab

Here, we implement an experimental design that combines a number of existing paradigms to investigate how systematic argument distinctions emerge in the evolution of novel manual communication systems. Pairs of participants take part in a silent gesture task, communicating using only gesture (no speech) about a set of events presented as a short discourse. The gestures each pair produce are used as the training 'language' for a new pair of participants, in an iterated learning design.

There is now a large body of experimental work using silent gesture, which has the benefit of constraining participants to use a modality that is not their primary one, reducing their ability to rely on existing linguistic knowledge. In addition, silent gesture experiments allow researchers to explore modality-specific constraints on communication. This method has been used widely to investigate the factors that affect the order of propositional arguments (Goldin-Meadow et al. 2008; Meir et al., 2014; Schouwstra and Swart, 2014). For example, Goldin-Meadow et al. (2008) found that speakers predominantly produced Agent-Patient-Action orders describe events when gesturing without speech, analogous to Subject-Verb-Object (SOV) order, irrespective of their native language. Further studies have highlighted the effect that iconic affordances of modality (Meir et al. 2014; Christensen, Fusaroli, and

Tylén 2016) or the semantics of the events (Hall, Mayberry, and Ferreira 2013; Schouwstra and Swart 2014) can have on the orders participants produce in the absence of prior language experience. However, as previously mentioned, sign languages do not rely exclusively on word order to denote relationships between predicates and arguments. In particular, the prevalence of spatial modulation and role-shift cross-linguistically suggests that the affordances of the manual modality affect which linguistic structures are used to denote those relationships. Intriguingly, there have been some anecdotal reports of interesting uses of space by participants in studies of word order. Both Gibson et al. (2013) and Hall et al. (2013) report some participants using space to disambiguate event arguments (e.g., using their body to signal the agent, and a gesture away from the body to signal the patient of an event). However, the focus of these studies was on how participants ordered their gestures, so examples of spatial modulation are uncommon.

Focusing on the use of space by silent gesturers, So et al. (2005) found that hearing participants used space to keep track of repeated referents when gesturing, suggesting that hearing participants with no knowledge of sign languages can use iconic space to disambiguate clause arguments across discourse in the gestural modality. Though we may infer from this study that spatial strategies might emerge right from the start in the evolution of a new language, a later study comparing the results from So et al. (2005) with Nicaraguan homesigners found differences between the silent gesturers who had an existing language model and the homesigners who did not (Coppola and So 2006). While the silent gesturers showed a higher use of spatial modulations overall than the homesigners, their productions were less 'language-like' than homesigners, with gesturers producing highly imagistic, pictorial gesture sequences that were unconstrained in *how* they use space, in comparison to homesigners. Therefore, silent gesturers do not spontaneously produce spatial modulations that reflect those found in sign language. Furthermore, emerging sign languages demonstrate increasing systematization and conventionalization of spatial systems across generations of the signing community (Padden et al. 2010; Meir 2012; Kocab et al. 2015; Montemurro et al. 2019). This suggests that an experimental model of the evolution of spatial modulation and role-shift should take language use and transmission to new learners into account, in addition to innovation by individuals.

We model both language use and transmission, respectively, with an experimental design that combines a director-matcher paradigm (Clark and Wilkes-Gibbs 1986; Garrod et al. 2007; Fay, Arbib, and Garrod 2013) and iterated learning (Kirby et al. 2015; Silvey, Kirby, and Smith 2015; Beckner, Pierrehumbert, and Hay 2017). Pairs of participants communicate with each other about events using only gesture and no speech. The gestures they produce are then used as the training language for a new pair of communicating participants. This combined design was used by Motamedi, Schouwstra, Smith, Culbertson, and Kirby (2019) among others (Winters, Kirby, and Smith 2014; Kirby et al. 2015; Silvey et al. 2015), demonstrating the evolution of both systematic and communicatively efficient languages in the manual modality.

Combining silent gesture with iterated learning in this way provides an experimental model of two defining processes in natural language: communication and transmission of language to new learners. This allows us to examine the cultural evolutionary mechanisms that enable the emergence of systematic, conventionalized constructions from initially nonlinguistic communicative signals. We extend this method here to test whether communication and transmission together lead to the systematization of forms, mediated by the iconic and bodily affordances of the modality, to represent complex predicate–argument relationships. In particular, we expect that our gesturers will rely on the iconic affordances of the modality, namely the potential of their own bodies to represent animate arguments. Further, we predict that these forms will become more systematized and conventionalized through use in communication and transmission to new learners.

2. Methods

2.1 Participants

Fifty participants were recruited from the University of Edinburgh's careers website. Participants were paid £7 to take part in the experiment, which took up to fifty minutes to complete. All participants were self-reported right-handed native English speakers with no knowledge of any sign languages. Participants who had taken part in previous similar experiments (e.g., a silent gesture task) were not allowed to participate.

2.2 Materials

Stimuli were orthographically presented *pairs* of events, designed to simulate a simple discourse (e.g., *Hannah is swimming. Sarah is walking.*). All pairs of events involved two actors, *Hannah* and *Sarah*, who in a particular event could either be the agent, the goal, or the end location of the event. The recurring actors in the experiment, and their distribution in the event pairs, were chosen to create enough ambiguity that participants had to signal both events in a pair to communicate effectively. For example, it was not the case that if Hannah was the agent of the first event that Sarah would always be the agent of the second, so participants could not avoid producing gestures for both events in a pair. In addition, we presented events as orthographic sentences, rather than as pictures to encourage participants to focus on the events as a whole rather than properties of individual agents and patients. By using event participants presented as orthographic names, we can signal that arguments are different without highlighting their individual properties (e.g., using distinctive features like hair, clothing or accessories). This is important, since the goal of this study is to understand how gesturers signal grammatical relationships, rather than how they represent individual arguments.

Verbs in the event pairs were from one of four categories, shown in Table 1. Spatial modulation in sign languages usually occurs with a subset of verbs that correspond to our nonphysical transfer category. Our verb sets represent categories for which both iconic spatial mappings and grammatical spatial mappings could be used. We selected verbs based on the types of verbs that fit in similar categories in natural sign languages, understanding that some of these concepts may be gestured iconically or through conventional gesture forms (such as wagging a finger for scolding). However, we did not expect access to common signaling strategies to affect how participants used space. Verbs in the experiment were all presented in the present progressive form. A given pair of events always used different verbs from the same category.

Table 1. List of verbs used in stimulus sentences in the experiment verbs are grouped according to each of the four verb types, and throughout the experiment verbs were presented in the present progressive form.

Description	Verbs
Verbs of motion, no specified end point	to cycle
	to run
	to swim
	to walk
Verbs of motion, specified end point, P	to cycle to P
	to run to P
	to swim to P
	to walk to P
Verbs denoting physical transfer of an object to a recipient, R	to kick a ball to R
	to give a book to R
	to send a letter to R
	to throw a hat to R
Verbs denoting nonphysical or metaphorical transfer to goal or recipient, R.	to help R
	to phone R
	to praise R
	to scold R

The use of spatial devices in sign languages often occurs across a stretch of discourse, tracking repeated referents (Fenlon et al. 2018; Lillo-Martin and Meier 2011). For this reason, we created sets, consisting of four pairs of events, one pair using each verb category. As detailed in the procedure section below, participants were trained on and communicated about events one set at a time. This allowed for the possibility of re-using gestures across trials to communicate about recurring discourse arguments. Figure 4 illustrates an example set of event pairs a participant might be exposed to in the experiment (a list of all event pairs used in the experiment can be found at https://osf.io/hp5md/).

Critically, two types of sets were used: same-agent sets, where the agent was held constant in each pair of events (e.g., Hannah was the agent of both events), and different-agent sets where the agent was different across paired events (e.g., Hannah was the agent in the first events in the pair, Sarah the agent in the second). The inclusion of both same-agent and different-agent sets created a pressure for participants to fully disambiguate event arguments to successfully communicate. In total, participants were trained on and communicated about four event sets—two same-agent sets and two different-agent sets.

2.3 Procedure

Pairs of participants, which we will refer to as *dyads*, were seated in separate experiment booths, each in front of an Apple Thunderbolt monitor with an affixed Logitech webcam. Monitor and webcam were connected to an Apple Macbook Air laptop running Psychopy (Peirce 2007), and VideoBox, custom software that allows streaming and video recording via networked computers (Kirby 2016). Dyads were organized into five chains of five generations, shown in Fig. 5. Participants from generations 2 to 5 took part in both training and testing stage. Participants in generation 1, the starting participants in each chain, only took part in the testing stage and had to improvise a way of communicating with each other.

2.3.1 Training stage

In generations 2–5, both participants in a dyad were first trained on gestures produced by a model participant in the previous generation of the same chain. The model was randomly selected from one of the two participants in the previous dyad, and participants were trained on all the gesture videos produced by the model. Both

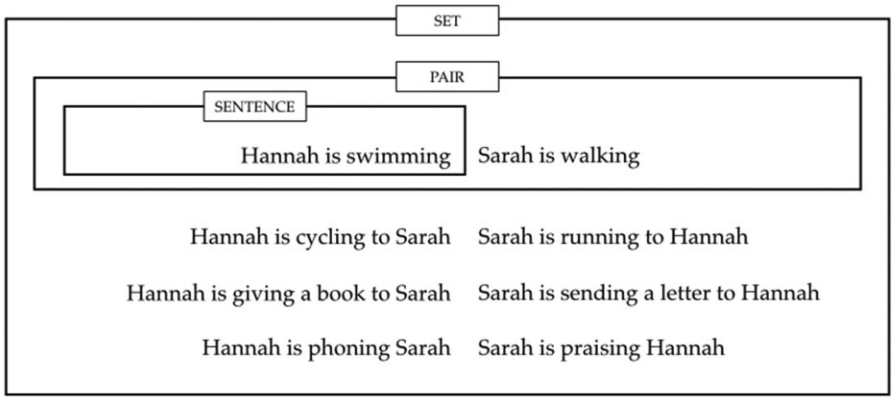


Figure 4. Examples of events used in the experiment, and how they are structured. At each trial, a pair of events is presented. Each pair uses two verbs of the same category, as given in Table 1. Pairs of events are grouped into sets of four, each pair exhibiting one verb type. The full body of events comprises four sets, two same-agent sets and two different-agent sets (given in Supplementary materials). This example is a different agent set because there are different agents for the two events in each pair.

participants in a dyad were trained on the same model, but each participant in the dyad was presented with the training videos in a different (random) order. Participants were given a three second countdown to prepare them for the start of each training trial, at which point a training video was shown on screen, with an array of four event pair choices, consisting of the target pair and three foil pairs. The three foils either differed from the target based on the verbs used or the agent configuration (either different-agent or same-agent), or on

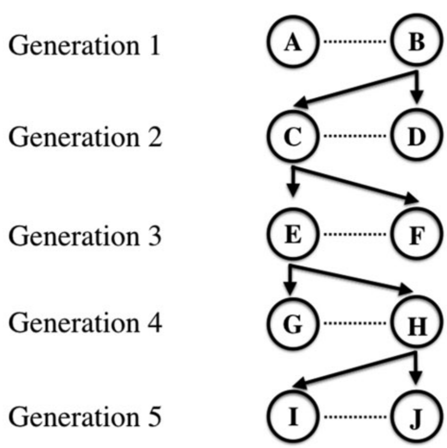


Figure 5. Transmission chain structure in the experiment. Solid lines with arrows represent transmission, dashed lines represent interaction. Pairs of participants (dyads) communicate with each other at each generation. Participants in generation 1 innovate gestures while communicating. Participants in generations 2–5 are trained on gestures produced by one of the two participants in the previous generation, before communicating with their partner.

both. An example is provided in Fig. 6. Foil verbs comprised the two remaining verbs in the same category and were randomly positioned in either the first or second event in a pair. Foil agent selection differs depending on the target agents. For same-agent targets, one (randomly chosen) event in the foil pair used the agent not present in the target. For different-agent target pairs, one of the agents (either Hannah or Sarah) was randomly selected to appear as the agent twice in the foil pair, so that participants had to communicate the event fully to disambiguate event arguments. Reflexive events, with the same individual mentioned twice, are never used in the experiment (e.g., *Hannah is praising Hannah* never appears). Crucially, in order to accurately communicate the target pair, a gesture sequence would need to fully disambiguate event arguments.

The array of matching choices was presented in a Psychopy window beneath the video stream, as shown in Fig. 7. Participants could make a guess at any point while watching the training video by pressing the number key corresponding to a pair of events. The video stopped streaming and a black window showed once a guess was made. Feedback was given after each guess, as shown in Fig. 7, presented for 5 s.

Participants completed a total of sixteen training trials, one for each target pair, organized into sets as described above. Participants took part in the training stage individually, and the order of presentation for event pairs was randomized for each participant.

2.3.2 Testing stage

During the testing stage, participants remained in individual experiment booths and interaction was facilitated

Target pair	Hannah is helping Sarah Sarah is scolding Hannah
Foil 1. Different verb from category	Hannah is **praising** Sarah Sarah is **phoning** Hannah
Foil 2. Different agent configuration	Hannah is helping Sarah **Hannah** is scolding **Sarah**
Foil 3. Both different	Hannah is **praising** Sarah **Hannah** is **phoning Sarah**

Figure 6. Example choices in array used for matching trials. Foils differed from the target in the verbs used (different verbs chosen randomly from the same verb category), in the configuration of the agents (e.g., same agent across events, instead of different agents), or both. Words in bold show differences between the target pair and the foil.

through video streaming between networked computers. Participants in the testing stage communicated with a partner using only gesture, taking turns to produce and interpret gestures in a director-matcher task (Clark and Wilkes-Gibbs 1986; Brennan and Clark 1996; Garrod et al. 2007). Participants both produced and interpreted gestures for all sixteen event pairs in the experiment, giving a total of thirty-two testing trials. Participants held the same role for four trials at a time, completing a full set of event pairs, before switching roles with their partner. Participants were notified of their role at the beginning of each set and at each trial. The order of event pairs was randomized within a set, and the order of sets for each directing participant was randomized.

As director, participants were presented with a pair of events and instructed to communicate both events to their partner, using only gesture. They were presented with a three second countdown at the beginning of the trial, and then shown the pair of events on screen for 5 s. They were presented with another three second countdown to prepare them for recording and streaming to their partner. When the recording and streaming started, participants saw themselves in the VideoBox window, with their image mirrored (the streamed feed remained unmirrored, see Fig. 7); the target event pair remained on screen. The director could stop the recording and streaming by pressing the space bar, upon which video streaming was terminated for both director and matcher. The director waited for the matcher to make a guess, before receiving feedback as shown in Fig. 7. The director did not see the full array of choices presented to the matcher.

As matcher, the participant had to interpret gestures produced by their partner. The matcher was given a three second countdown at the start of the trial, and waited while their partner was shown the target events. Another three second countdown prepared the participant for streaming from their partner. The matcher saw their partner gesturing on screen, unmirrored, and was presented with an array of four event pairs, and matching proceeded as in the training stage (shown in Fig. 7). The experimental design allowed for interruption, so that either the director could terminate recording and streaming by pressing space bar, or the matcher could do so by making a selection. As in training, feedback indicated the selection and the target.

3. Results

3.1 Gesture coding scheme

Gestures produced by participants were coded for the presence of agent, patient/goal, and verb. The presence of an agent or patient/goal was coded if any agent or patient/goal could be inferred from a gesture, such that the coder could explain which part of the sequence identified the agent or patient/goal. Gestures pertaining to neither agent nor goal were coded as part of the verb form. If an agent, patient/goal or verb was present in a trial gesture, we coded the type of gesture used to represent each element. For agents and patients, gestures were coded either as specific handshape (e.g., holding up 1 finger), using the whole body or torso, or using indexed locations around the gesturer. For verbs, we coded gestures as handling gestures (where the handshape represents a manipulation of an object), instrument gestures (where the hand represents an object), body gestures (where the gesturer performs an action with the whole body or torso, such as running), and descriptor gestures (where the hands outline the size or shape of an object). Finally, we coded symbolic verb gestures, where there was a gesture as part of the verb sequence, but where the meaning of the gesture was not transparent to the coder. These account for 5%

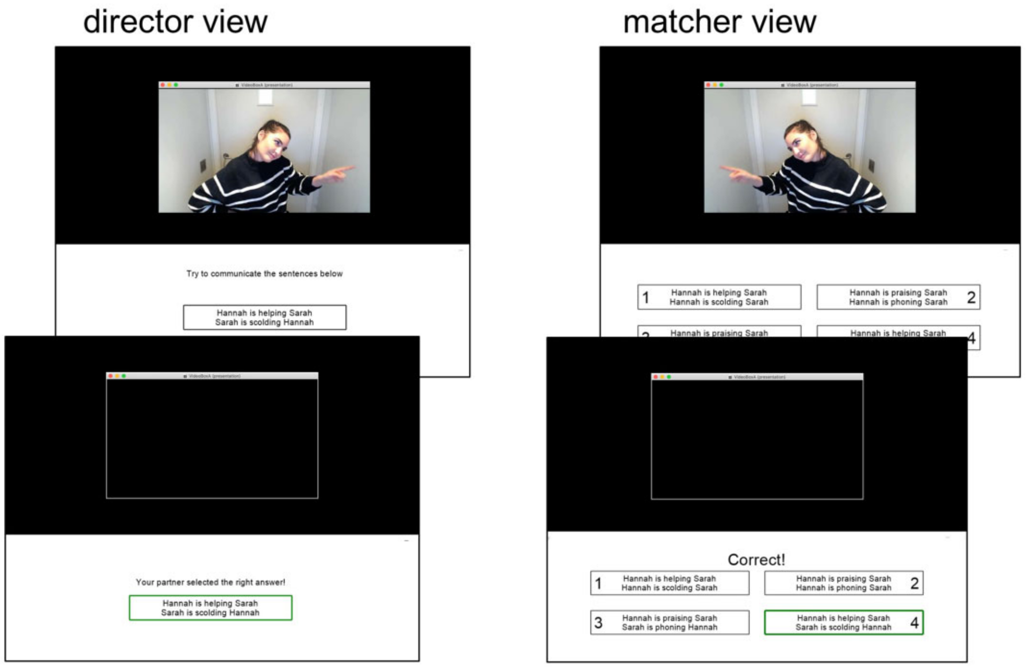


Figure 7. Testing procedure: example of a testing trial from the director view (left) and matcher view (right). During recording, the director saw the target event pair on screen as well as a mirrored video stream of themselves and had to produce gestures to communicate both events in a pair. The matcher saw the unmirrored video stream of the director producing gestures communicating the event pair. Underneath the video, they saw an array of event pairs, including the target pair and three foil pairs. Both participants were given feedback once the matcher had made their selection.

of all verb gestures coded. Videos from two trials could not be coded due to a technical error in recording.

In addition to gesture types, gestures were coded according to positional parameters: location of a gesture, orientation, and direction of movement were coded, and we noted whether the gesture was produced in a neutral (directly in front of the signer) or non-neutral (away from the body) location. Finally, we noted whether or not a separate verb path was present (i.e., a movement path articulated separately from the verb gesture), and if so the direction of this path. Gestures were coded by the first author. In a first pass, another coder, who was blind to the hypotheses of the experiment, coded 20% of gestures produced for a target event. Analysis of inter-coder agreement revealed low agreement on the parameters concerning location and position (neutral/non-neutral). As such, these parameters were re-coded by the first author based on an amended coding scheme and 20% of the data (all parameters) were coded again by another coder blind to the hypotheses of the experiment. The median percentage agreement between coders from this second pass was 86.40% across 10 variables (range 74.69–99.06%).[1]

3.2 Qualitative results: strategies for argument distinctions

Our analysis focuses on how participants distinguish between agents in event pairs. There are a number of ways in which participants create these distinctions, which largely map onto three main strategies that we term the lexical, body, and indexing strategy, respectively. Below we describe each strategy and illustrate how it emerged and became systematic across generations of a given chain. Links to videos for all examples shown here can be found at https://osf.io/hp5md/.

3.2.1 Lexical strategies

Lexical strategies, used by chains 1 and 5, are characterized by the type of gesture used to denote an agent. In both chains in which this strategy is primarily used, participants vary the agent gesture based on handshape. Specifically, a 1-handshape denotes the first agent, and the 2-handshape denotes the second agent in the event pair (shown in Fig. 8).

These gestures are initially used to simply distinguish between the first and second event of a target pair. In this case, the 1-handshape signals the first event in the pair, and does not refer to any particular argument in that production. Initially, they may be considered as numerical markers akin to list buoys used in natural sign languages that enumerate a set of concepts (Liddell 2003). However, examples from each chain suggest that these forms evolve to distinguish between agents rather than events, and stand in for specific agents rather than enumerate them. Figure 9 illustrates how these forms are used as agent markers rather than discourse markers signaling the order of events. In Fig. 9a, showing an example from chain 1, the participant reuses the 1-handshape in gesture sequences for both events, signaling that the agent is the same in both cases. In Fig. 9b, the 1 and 2 handshapes stand in for agent and goal, respectively, both gestures referring to participants in the same target event.

3.2.2 Body strategies

Body strategies, used by chains 2 and 4, involve the use of the participant's body orientation to signal differences between arguments across event pairs. Figure 10 illustrates the use of body orientation to denote the agents in an event pair; the participant in this example shifts the orientation of her whole body to distinguish between Hannah and Sarah in target events. Some participants using this strategy vary the orientation depending on whether the event pair is a same-agent or different-agent pair, disambiguating the context but not particular agents. However, in the example in Fig. 10, the participant not only varies their body position to denote that the agents in the two events of a pair are different, but differentiates arguments consistently, gesturing actions by Sarah in a right-oriented position, and actions by Hannah in a left-oriented position. This general strategy is iconic, and relies on the similarity of the gesturer's body to other animate bodies (Meir et al. 2014). This phenomenon of 'body-as-subject' is commonly attested in sign languages; the signer's body represents the subject, or highest ranking thematic role of a proposition (Meir et al. 2007), and the object/patient may be left unspecified.

3.2.3 Index strategies

One chain in the experiment (chain 3) uses a strategy that relies mainly on indexing. Participants differentiate agents in an event pair by indexing locations in the gesture space to represent them, primarily with a deictic point (see Fig. 11). Indexed locations are usually opposite to each other, but this opposition can be set up on the z-axis, where one locus is the gesturer's own body (Fig. 11a), or on the x-axis, where both loci are set up parallel to the signer's body (Fig. 11b). While indexing in chain 3 is initially innovated along the z-axis, it develops through

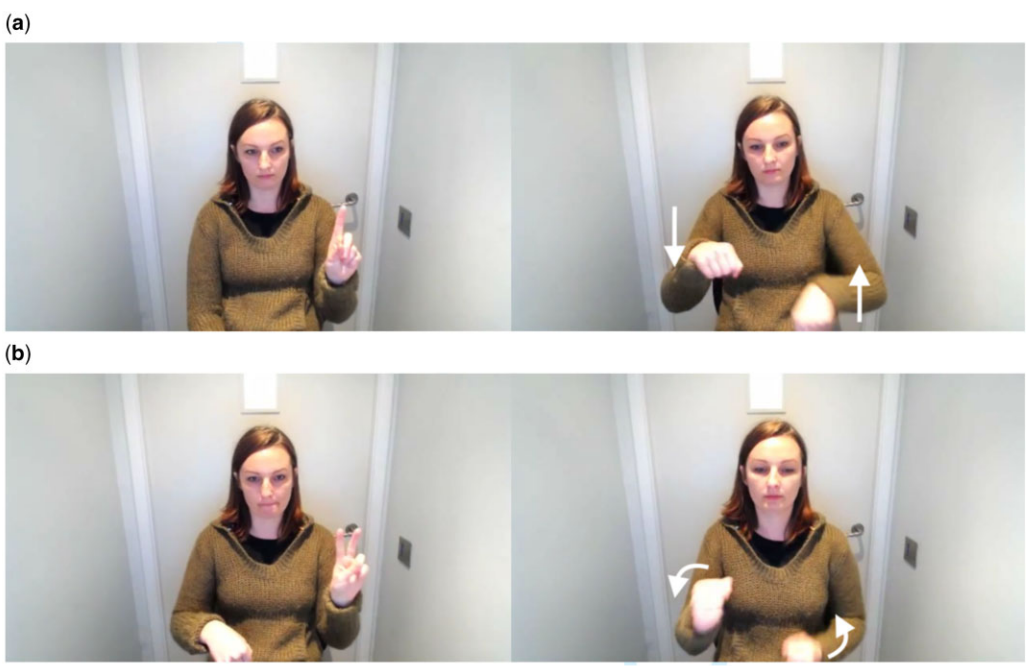
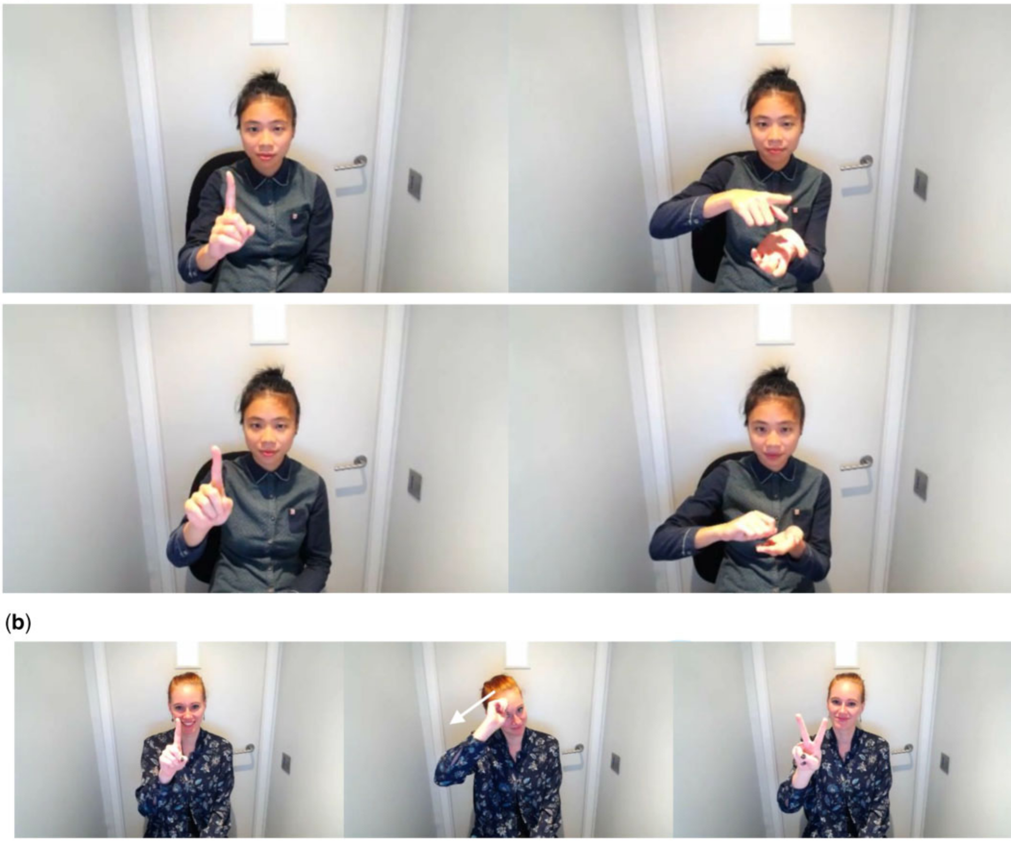


Figure 8. Use of the lexical strategy. The participant represents different agents in a different-agent event pair using one to denote Sarah and two to denote Hannah. (**a**) Sarah is cycling. (**b**) Hannah is running. Example from chain 5, generation 3. Links to corresponding videos for all images shown in this section can be found at https://osf.io/hp5md/.

(a)



(b)

Figure 9. Examples showing the use of one and two gestures to denote agents. In (**a**), the participant uses the one gesture for both agents in a same-agent trial; here the gesture denotes the agent rather than the sequence of events. In (**b**), the participant uses both one and two gestures in a gesture sequence, where one notes the agent and two marks the goal. Gesture examples are taken from generation 5 participants of chain 1 (**a**) and chain 5 (**b**). Sarah is sending a letter to Hannah and Hannah is giving a book to Sarah. Hannah is throwing a hat to Sarah and Sarah kicking a ball to Hannah.

generations to be enacted on the *x*-axis, consistent with findings from naturally emerging sign languages (Padden et al. 2010; Montemurro et al. 2019).

The indexing strategy allows the possibility of modulating the verb gesture with respect to the location of the agent or goal. For instance, Fig. 12b shows a gesture in which the predicate gesture for *giving a book* is directed from the location of the agent to the location in which the goal is indexed. This contrasts with the gesture shown in Fig. 12a, where the direction of movement for the action (here *throwing a hat*) signals no relationship between the locations indexed for either agent or goal.

3.3 Quantitative results: the evolution of signaling strategies

The measures described below are based on comparing the agents[2] in the first and second event of a pair, and noting whether they were expressed differently (e.g., using one of the strategies described above). Note that creating systematic distinctions between agents in a target event pair is not the only way for participants to complete this task. Participants' gestures could encode no difference in the way agents are marked (e.g., body gestures in neutral position used for all trials). Alternatively, participants could innovate gestures that represent events holistically. However, recall that our experiment was designed such that some trials featured the same agent across pairs in a set, whereas others featured different agents across pairs in a set. If participants are using gesture to distinguish between agents in these mini-discourses, then we expect to see distinct patterns of behavior across these set types. Specifically, gestures in different-agent sets should create distinctions between agents.

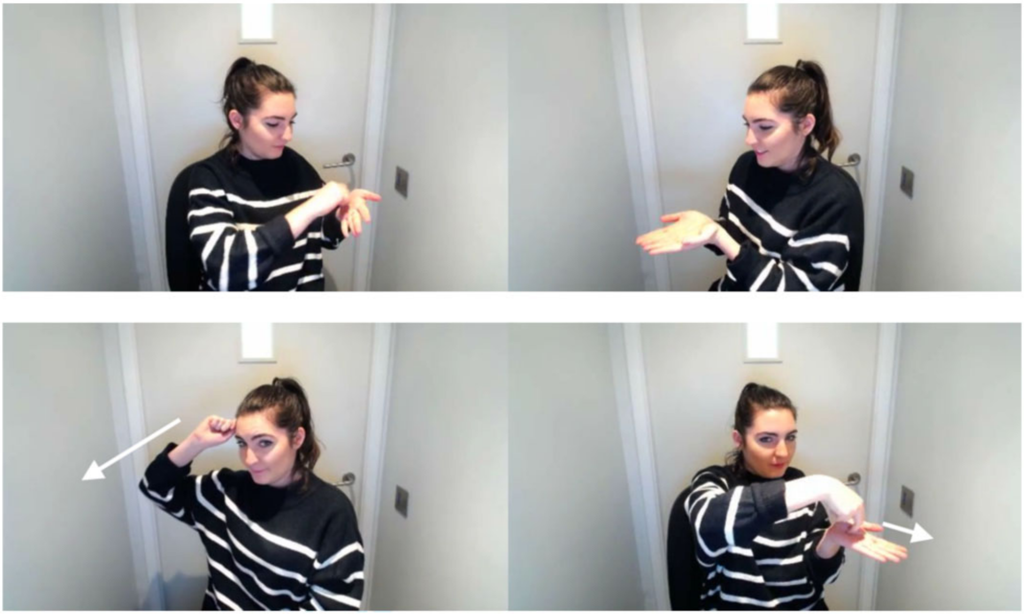


Figure 10. Use of the body strategy. In both event pairs, the participant demonstrates the difference between agents by body orientation. Furthermore, body position denotes particular agents; orientation to the right signals Sarah, orientation to the left signals Hannah. These examples were produced by a generation 5 participant in chain 2. (**a**) Sarah is helping Hannah. (**b**) Sarah is helping Hannah.

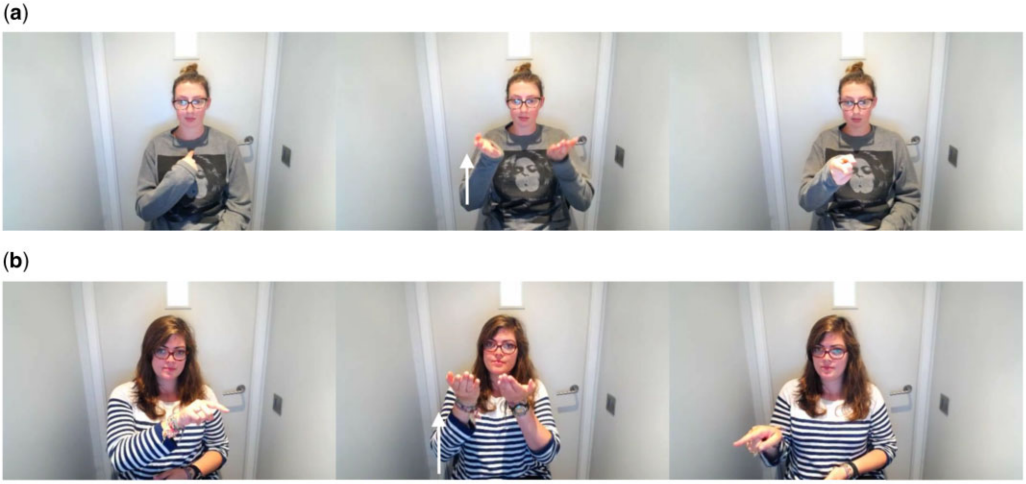
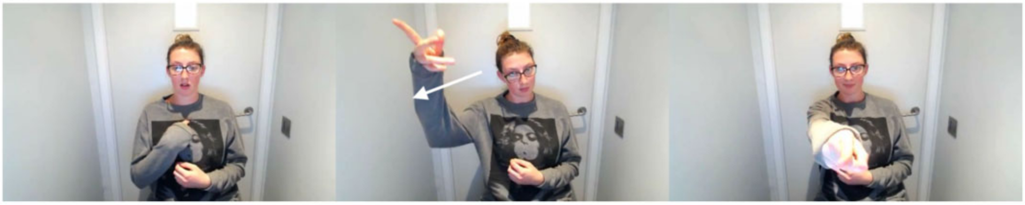


Figure 11. Examples of the indexing strategy, where separate locations are indexed to represent different arguments from the target event: (**a**) shows a participant using indexed locations on the *z*-axis. In the first frame, she indicates the agent by pointing at her own body; in the final frame a point outward signals the goal of the event and (**b**) shows a later generation, where the reference axis has moved, and predicate arguments are indexed on the *x*-axis. Both agent and recipient are now set up in opposing locations parallel to the gesturer's body. The participants shown here participated in chain 3, generation 2 (**a**) and generation 3 (**b**). (**a**) Hannah is throwing a hat to Sarah. (**b**) Hannah is giving a book to Sarah.

3.3.1 Differentiated agents across trial types

Figure 13a illustrates distinctions between agents in event pairs within same-agent and different-agent contexts. We plot separately the four possible ways of creating such distinctions based on the parameters in our coding scheme: varying the explicit agent gesture (shown as agent type in Fig. 13a), varying the location of the agent gesture, varying the location of the verb

(a)



(b)

Figure 12. Examples of indexing gestures where the verb path is neutral with respect to indexed locations (**a**) and where the verb path moves between two indexed locations (**b**). In (**a**) the path of the verb does not move between location of the indices denoting Hannah and Sarah. In (**b**) the verb gesture moves between the location of the agent and the goal of the target event, Sarah. Examples here are from participants from chain 3, in generation 2 (**a**) and generation 4 (**b**).

gesture, and varying the direction of the verb path. Importantly, the ways in which these parameters vary map onto our classifications of strategies outlined in Section 3.2. For example, lexical strategies primarily vary the agent gesture itself to mark a distinction between agents, and indexing strategies primarily vary the agent location. The body strategy, where agent and verb are gestured simultaneously necessarily varies agent location, verb location and verb path simultaneously, and therefore we see distinctions made based on each of these parameters separately. The data illustrated here confirm that chains converge on particular strategies. Chains 1 and 5 are described above as using a lexical strategy; here, we see that distinctions between same- and different-agent gestures are primarily made based on the agent type. Chain 3, which was described as using an indexing strategy, primarily shows differences between event types based on the location of the agent gesture. Finally, chains 2 and 4 were described as using a body strategy, where they change the orientation of their body position to communication agent distinctions. In Fig. 13a, we see that both chains show differences between same- and different-agent events based on three parameters, agent location, verb location, and verb path, which vary simultaneously to signal different agents.

We can also focus on whether distinctions between agents evolve in the same way, regardless of the particular strategy a chain uses. For this, we scored each trial with a binary measure: is a difference signaled between agents in both events of a target pair (i.e., using any feature shown in Fig. 13a) or not? Using this measure (shown in Fig. 13b), we analyzed the proportion of agent differentiation using a logistic mixed effects regression, implemented with the lme4 library (Bates et al. 2015) in R (R Core Team 2008). The model included event type (with same-agent type as the intercept) and generation (with generation 1 as the model intercept) as fixed effects, as well as their interaction. We included chain, target pair, and participant as random effects with random intercepts, with the random effects structure for participant nested within chains. A by-chain random slope did not allow convergence. As such, we report results here from a model without the random slopes.[33] The model demonstrated a significant effect of event type (different-agent context: $\beta = 3.05, \mathrm{SE} = 0.55, z = 5.69, P < 0.001$), indicating that participants gesture differently based on whether agents in the event pair are the same or different from the first generation. We found no significant effect of generation for the same-agent event type ($\beta = -0.006, \mathrm{SE} = 0.09, z = -0.06, P = 0.95$). However, the model demonstrated a significant interaction between event type and generation ($\beta = 0.40$, $\mathrm{SE} = 0.17$, $z = 2.30, P < 0.02$), indicating that the frequency of agent differentiation increased over generations in different-agent contexts. This pattern of results shows that the effect of generation differs for the two event types: differentiation of agents is significantly more likely in different-agent contexts versus same-agent contexts, and this divergence increased over generations. In other words, although chains of participants converge on different strategies, all chains eventually differentiate

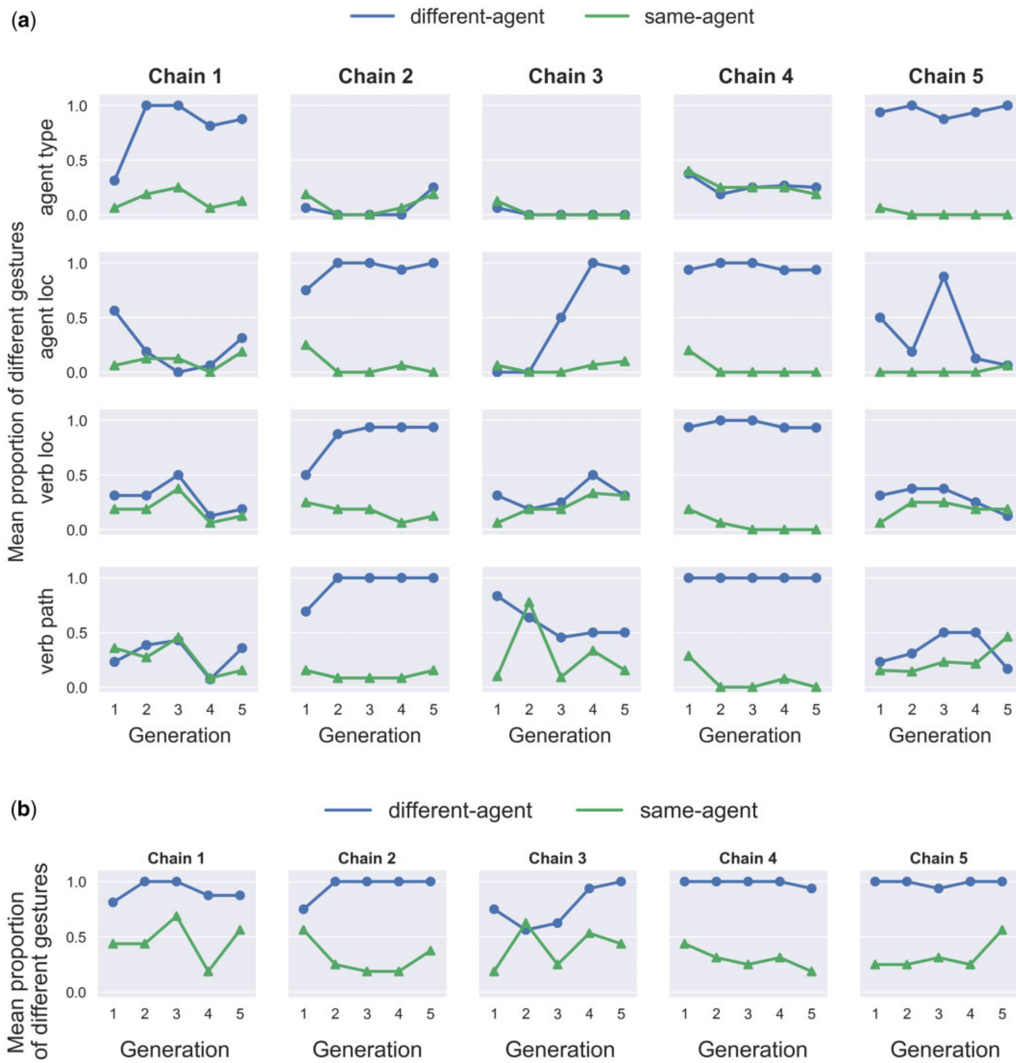


Figure 13. Proportion of gestures that differentiate agents in the target event, in same-agent (green lines) and different-agent (blue lines) contexts: (**a**) shows these differences based on which coded parameter varies (agent type, agent location, verb location, verb path), and columns show the values for each chain, at each generation; (**b**) collapses the difference measure to a binary variable—is there a difference between events within a pair or not? All chains show differences based on the context in which the agents occur, demonstrating a higher proportion of differentiation in the different-agent context. However, chains differ in the gestural parameters used to make these distinctions, which map onto the descriptive strategies outlined in Section 3.2.

between agents in contexts where potential ambiguities arise.

We also assessed whether the distinctions participants make affect communicative accuracy—that is, whether their partner can accurately identify the correct target pair from the matching array. We calculated the proportion of correct responses from matching participants for each pair at each generation. Accuracy was high across all chains and generations ($M = 0.9$, $SD = 0.3$, range = 0.68–1.00). To analyze change in accuracy over generations, we ran a logistic mixed effects model on the binary variable of accuracy (1: correct, 0: incorrect), with a fixed effect of generation. We included random intercepts for participant and target pair, and the random intercept for participant was nested in chains. The model did not reveal a significant effect of generation ($\beta = 0.23, \mathrm{SE} = 0.15, z = 1.60, P = 0.11$), suggesting that, while accuracy is high, it does not change significantly over generations. As Fig. 14 illustrates, we do not see clear differences in accuracy based on the strategy used by participants. Analysis of the errors participants make indicate that a large majority (~89%) involve cases where participants select the wrong agent configuration.

3.3.2 Use of non-neutral locations across strategies

Here we focus on when participants produce gestures in neutral and non-neutral locations. While the use of space can correlate with the strategies participants use, it does not necessarily do so. Therefore, looking at use of space alone is revealing; placing referent signs outside neutral signing locations is often indicative of grammatical spatial modulation in natural sign languages (Senghas and Coppola 2001). The proportion of gestures performed in non-neutral locations is shown in Fig. 15.

Figure 15 indicates that non-neutral gesture locations are used to some extent by all chains. However, here we restrict ourselves to a qualitative analysis, since the use of non-neutral locations differs dramatically across chains depending on the differentiation strategy they use. Chains 1 and 5, which use the lexical strategy, exhibit a reduction in the use of non-neutral locations, which become redundant as specific gesture forms are conventionalized. In the remaining chains, participants use non-neutral gesture locations to differentiate agents, either across all contexts (chains 3 and 4) or in different-agent contexts only (chain 2). Interestingly, non-neutral space is not used in contrast to neutral space to differentiate arguments. Rather, the contrast relies specifically on that chain's differentiation strategy. For instance, participants in chain 3 demonstrate a contrast between different-agent and same-agent contexts by using different indices for different agents, but do not contrast the use of neutral and non-neutral locations; all indices are placed in non-neutral locations. Use of non-neutral gesture locations emerges early to denote agents that differ from each other, as the participant's body in neutral position cannot contrast multiple agents.

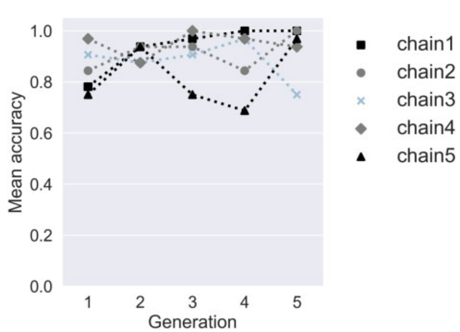


Figure 14. Proportion of trials accurately identified by the matching participant. Individual lines and markers show each chain, common colors indicate common strategies used (black: lexical, gray: body, blue: indexing). Accuracy is high and does not change significantly over generations. We do not see clear differences in accuracy based on the referential strategy used.

3.3.3 Distinction strategies based on verb type

Finally, we analyzed the proportion of gestures that differentiate agents in an event pair, based on the four verb types present in the target events (see Fig. 16).

A logistic mixed effects analysis investigated the effect of verb type on the proportion of differentiated agent gestures in an event pair. Verb type was included as a centered fixed effect along with generation and their interaction. Chain, target event, and participant were included as random effects with random intercepts, with the random intercept for participant nested within chains. As before, we included a by-chain random slope of generation, as well as a by-target slope of verb category. The random slope model did not allow convergence, so we report analysis based on the model without random slopes. Our model found no effect of verb type ($\beta = 0.50, \mathrm{SE} = 0.41, z = 0.1.23, P = 0.22$) or generation ($\beta = 0.09, \mathrm{SE} = 0.08, z = 1.15, P = 0.25$), nor a significant interaction between the two ($\beta = -0.03, \mathrm{SE} = 0.06, z = -0.54, P = 0.59$).

The results from the model suggest that participants do not systematically vary how they differentiate agents in target events pairs, based on the verbs that represent those events. This result differs from many older sign languages, such as ASL, where the use of spatial modulation occurs with a subset of predicates denoting non-physical transfer. However, these results are consistent with data from an emerging sign language, ABSL, which shows no difference in the use of spatial modulation based on verb type (Padden et al. 2010).

3.4 Results summary

Participants in our experiment produce gestures that differentiate between agents across sets of target events. Chains of participants that interact with and learn from each other converge on strategies to create these distinctions. While many of these strategies rely on the iconic affordances of the manual–visual modality, these strategies conventionalize over generations. Participants demonstrate contrasts between discourse contexts of same-agent and different-agent pairs, producing more differentiated forms in different-agent contexts than in same-agent contexts. While differentiating forms often involves using non-neutral space, the contrast between neutral and non-neutral locations does not itself serve to distinguish between agents in most chains. Finally, our findings suggest that strategies for distinguishing agents

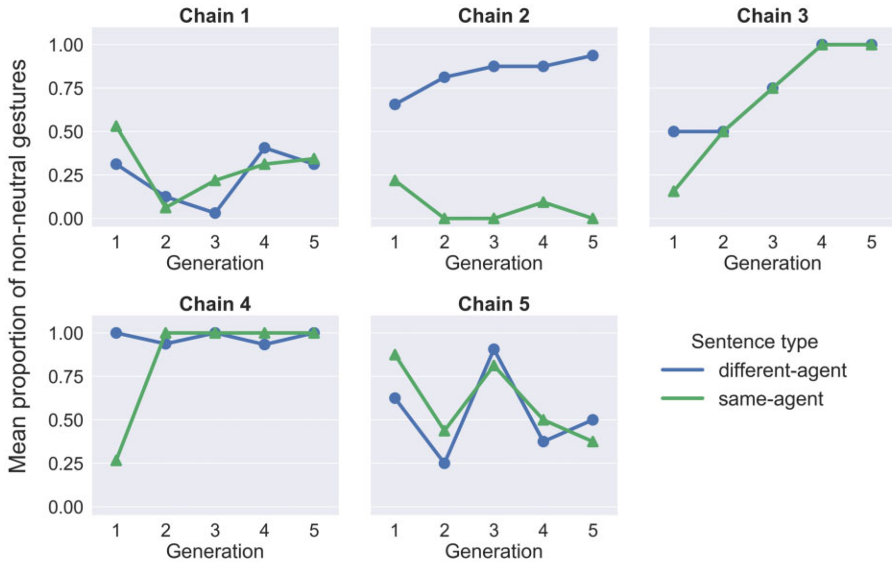


Figure 15. Proportion of gestures performed in non-neutral locations, in different-agent and same-agent contexts. Blue lines represent different-agent contexts, green lines same-agent contexts. Individual plots are shown for each chain, over all generations of the chain.

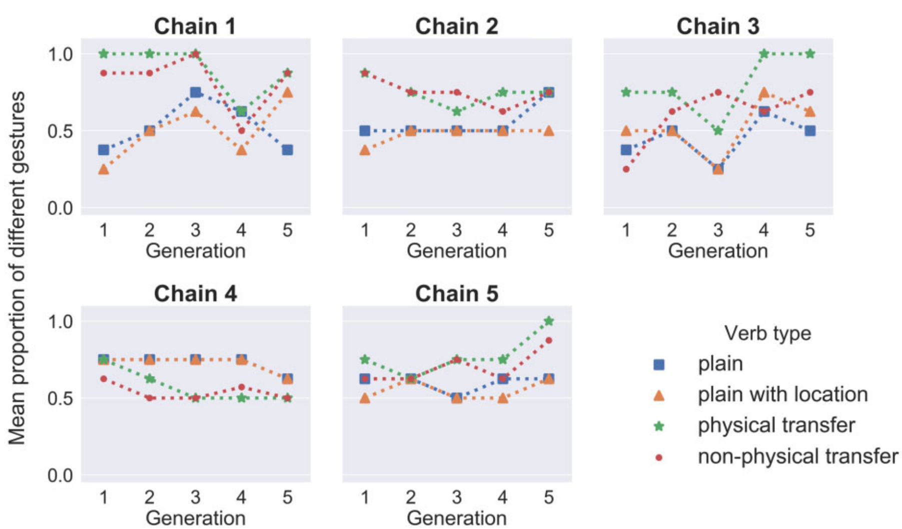


Figure 16. Proportion of gestures differentiating agents in target events by verb type. Individual plots are shown for each chain, over all generations of the chain. Colored lines show different verb types.

across events did not differ based on the verb type. Rather, these strategies were used across the whole system of gestures.

4. Discussion

The primary aim of the present study was to investigate the development of strategies to signal predicate–argument relations in manual communication systems, focusing on the use of iconic, body-centered strategies such as spatial reference. Here, we discuss the strategies that participants use to differentiate between agents in the target events and how they relate to distinctions in natural sign languages, before discussing what our experimental results suggest about how these distinctions evolve.

4.1 Differentiation mechanisms in natural sign languages

Participants in the present study demonstrated the use of space in their gestures from the first generations of transmission chains, though not all chains converged on systematic spatial strategies. The lexical strategies used by chains 1 and 5 do not make contrastive use of space, but contrast agents across target events using distinct gestural forms. Participants used conventional number handshapes, which initially indicate the sequence of events (i.e., events 1 and 2). However, these gestures systematize over generations to stand in for individual names, or pronoun-like forms for the two recurring agents. It is unsurprising that some participants relied on lexical contrasts; all languages use lexical forms to distinguish people or objects. In the context of our experiment, use of these lexical forms and the use of space appear to be non-overlapping; no chain develops a strategy that redundantly contrasts both ‘lexical’ form and spatial location.

The remaining chains make use of space to distinguish between agents. The body strategies exemplified by chains 2 and 4 involve movement of the participant’s body to positions that distinguish between agents; for the indexing strategy used by chain 3, participants’ bodies remain in neutral position, but indices referencing agents in the target events contrast in position. The indexing strategy could be considered the closest parallel to the spatial modulation seen in sign languages, where deictic points index referents in space and the paths of agreement verbs move between indexed referents, as shown in the BSL example in Fig. 1. However, modulating body position to signal role-shift is also well-attested in sign languages (Padden 1990; Cormier et al. 2015; Kocab et al. 2015; Fenlon et al. 2018), as shown in Fig. 2. Role-shift is used to differentiate predicate arguments (Padden, 1990), and the use of the body to represent animate agents in both contexts indicates that these mechanisms are not unrelated to each other (Cormier et al. 2015). Furthermore, use of the body to represent the agent, the highest thematic role in a proposition, is proposed by Meir et al. (2007) to hold a privileged position in sign languages. Across all chains in the experiment, participants show iconic use of the body to represent predicates in target events. We discuss the iconicity of such gestures in Section 4.2.

The development of the index strategy in chain 3 mirrors the evolution of spatial modulation observed in two naturally emerging sign languages. Participants in generation 1 index referents on the z-axis, pointing at themselves for the agent, and indexing the space in front of them for the goal of the target event. By generation 3, however, indexing has been abstracted away from the body to the horizontal x-axis. This change is also seen in the evolutionary trajectory of two young sign languages, NSL and ABSL (Padden et al. 2010; Flaherty 2014; Montemurro et al. 2019). Signers in older generations show a preference for verb paths that are oriented with respect to their own bodies, on the z-axis, while signers from more recent generations show an increase in the use of abstracted verb paths on the x-axis. Padden et al. (2010) and Meir et al. (2007) suggest that the two axes represent competing iconicities: one which represents animate agents with the most accessible resource available, the signer’s body, and one which is able to iconically contrast non-first person agents, but which requires abstraction away from the body.

4.2 Iconic representation through use of the body

All participants rely on iconic use of the body to represent agents and agentive actions in the target events. The body has been suggested to be a fundamental device for representation in sign languages (Taub 2001; Meir et al. 2007); for example, forms that frame an event in relation to the body (i.e., using the z-axis) appear before forms that abstract away from the body (Meir et al. 2007; Padden et al. 2010). Even in older sign languages such as BSL and ASL, signers show a preference for body-situated verb paths (Cormier et al. 2015), and there are particular subsets of verbs that favor iconic representation, such that sign forms are anchored to the body, with the path from agent to goal always originating from the signer’s body (Liddell 2003; Meir et al. 2007). As such, the prevalence of body-use in the present experiment supports the primacy of the body for representing animate agents in the manual–visual modality.

We included four different verb types to test whether different semantic relations lead to different gestural representations. For example, it may be that spatial verbs serve as a locus of change in a spatial agreement system because they are grounded in real-world spatial relationships and therefore lend themselves to iconic spatial mappings. In fact, we found that participants’ gestures did not demonstrate any differences between verb types, with similar strategies used to represent agents across all four categories. The absence of variation between verb types is consistent with data from ABSL and ISL (Padden et al. 2010), indicating that signers in two young sign languages use similar strategies independent of verb type, particularly in early generations of both languages. This contrasts with the linguistic systems of older sign languages, like ASL and BSL, in which verbal spatial modulation usually occurs with a subset of verbs (Sutton-Spence and Woll 1999; Lillo-Martin and Meier 2011). As such, it is possible that our data recapitulate the very early emergence of systematic spatial reference in manual communication systems that could further develop semantic differences over longer timescales.

The gestures produced throughout the experiment rely on iconicity to signal agents and actions, consistent with spatial systems in natural sign languages (Cormier et al. 2015; Schembri et al. 2018). However, we also see changes in *how* participants use iconicity. For example, the movement in chain 3 from body-situated gestures to indices placed on the x-axis suggests abstraction away from the body. Participants' initial preferences for situating their own body as the agent gives way to representations where third-person referents can be gestured as separate from the gesturer. Previous research in experimental contexts has suggested that iconicity sits in opposition to systematic structure (which must allow for some generalization across linguistic form to signal similarities in meanings), and thus that iconicity reduces as systematicity is built (Theisen, Oberlander, and Kirby 2010; Theisen-White, Kirby, and Oberlander 2011; Verhoef, Kirby, and Padden 2011; Roberts et al. 2015). Under this account, we would expect iconicity to reduce as spatial reference systems become more systematic. There may be some evidence of this in our data. For example, participants position arguments in maximally contrastive positions, such as left versus right, or close to the gesturer's body contrasted with the furthest visible position away from the body on the z-axis. Maximal contrasts, though still iconic, do not provide a direct iconic mapping from target meaning to gesture, as the maximal space between gestures does not convey differences in the space between Hannah and Sarah in the target events. As such, the use of maximal contrasts indicates the systematization of spatial mappings to represent the thematic roles of agent and goal, rather than direct spatial relations between Hannah and Sarah.

However, understanding these findings in relation to iconicity in natural sign languages is more complex. While older sign languages such as ASL may show a greater prevalence of arbitrary spatial modulations than the young sign languages that have been observed (Padden et al. 2010), arbitrary spatial mappings are in no way obligatory, and in fact, where possible, signers tend to use iconic spatial mappings to represent grammatical relationships (Liddell 2003; Cormier et al. 2015). If referents are present in relation to the signer, the signer will often use real-word spatial relations as the basis for the spatial mapping, but even in the absence of the referents themselves, signers can use iconic spatial features in their representations, such as representing differences in height between arguments (Liddell 2003). Furthermore, spatial reference on the z-axis is still common in languages such as BSL (Cormier et al. 2015), suggesting that, even if reference on the x-axis becomes more common over time, it does not displace z-axis use by any means, but is used selectively. Finally, it is also possible that an abstraction away from the body is not necessarily a reduction in iconicity. That is, a representation on the x-axis requires abstraction from the body, such that an animate agent is not represented by an animate body, but that mapping may in fact be more iconic than a representation on the z-axis, especially if it reflects a real-world spatial relationship. In this case too, the body's ability to represent any animate agent and the first-person agent (the signer themself) may be in conflict with each other (Meir et al. 2007).

We assert then, that the gestures participants produce demonstrate an early reliance on the body, with participants using the body iconically to represent both agents and agentive actions. In some cases, particularly as shown by our findings from chain 3, we see a gradual abstraction of spatial reference away from the gesturers' bodies over generations, supporting evidence from naturally emerging sign languages. However, participants continue to rely on iconic representations throughout generations, exploiting the affordances of their bodies to represent animate agents, and changing how they exploit these affordances to create distinctions between event arguments.

4.3 The evolutionary pathway for argument structure and the limits of space

Evidence from natural sign languages suggests that the systematic use of space does not emerge fully formed in a language, but takes time to develop (Meir et al. 2007; Padden et al. 2010; Lillo-Martin and Meier 2011). We have applied experimental methods to this question to understand the conditions under which spatial modulation systems might emerge. Our results demonstrate that nonsigning participants use spatial modulation to distinguish between event arguments, consistent with previous research (So et al. 2005), but can make use of other resources afforded by the modality to signal relations between arguments, as do natural languages. Though use of the body provides a natural starting point for representing animate arguments, it is not the only resource for creating systematic distinctions. Indeed, some sign languages do not appear to have spatial reference systems at all (Nyst 2007; Vos 2012).

Perhaps, surprisingly, we do not find that the changes participants implement over generations lead to a cumulative increase in communicative accuracy; rather, matching accuracy is high from the first generation and does not change significantly over generations. We do not see a clear difference in communicative accuracy based on the strategy participants' use, with all strategies facilitating successful communication. However, we suggest that communication between interacting participants does play a role in how participants construct their gestural systems. Participants in the study created distinctions that demonstrate an adaptation to the communicative context (Fedzechkina, Jaeger, and Newport 2012; Winters et al. 2014)—argument distinctions evolved in cases where it was necessary to distinguish between different agents and less often in cases where the agent was the same. In sign languages, signers can flexibly exploit the iconic affordances of the body to fit the communicative context. Meir et al. (2013) describe one such scenario, in which the role of the body as first person competes with the role of body as an agent which is not necessarily first person (e.g., 'I brushed my hair' compared with 'My mother brushed my hair'). They describe that, where the default interpretation of forms using the body (with the signer 'brushing' their own hair) would be reflexive, in ISL, the conflict between body as first person and body as agent is frequently resolved by using spatial modulation, signaling the agent as separate from the signer's body (e.g., by signing the act of brushing on the body, and then signaling a point away from the signer to indicate a nonfirst-person agent). Interestingly, this adapted pattern is not preferred in ABSL, which shows a preference for body-anchored forms even in the conflicting context.

In addition to the adaptation to communicative context that chains settle on distinct strategy points to an increase in communicative efficiency. While at generation 1, chains create distinctions on multiple parameters that mix the lexical, indexing and body strategies to different extents, by generation 5, each chain can be characterized as primarily conforming to one of those three strategies. As such, the systems participants produce are communicatively efficient, achieving high accuracy while eliminating the effort that comes through articulating multiple redundant strategies (Gibson et al. 2019).

Furthermore, we do not expect communicative efficiency to be the only pressure acting on the systems participants learn and produce. Previous experimental research has investigated the roles that both learning and communication play in the evolution of novel communication systems (Kirby et al. 2015; Smith et al. 2016; Nölle et al. 2018; Motamedi et al. 2019; Raviv, Meyer, and Lev-Ari 2019). For example, Motamedi et al. (2019) analyzed the emergence of structure in novel gestural communication systems, comparing a transmission + interaction condition where participants both learned gestures and communicated with them, to conditions with only transmission or only repeated interaction. Their findings showed that only the combined condition led to forms that were both systematic and communicatively efficient. In comparison, the gestures produced in the interaction-only condition were communicatively efficient but not systematic; in the transmission-only condition, when learning occurred without communication, gestures were systematic but showed high levels of redundancy. As such, our results align with those of Motamedi et al. (2019). Participants demonstrate communicative efficiency, with redundant strategies being lost but communicative accuracy maintained. We also see changes over generations, with the way participants signal event arguments across different- and same-agent event types becoming increasingly distinct over generations. Interestingly, the strategies that participants use appear to change at different rates over generations, with the indexing strategy used by chain 3 showing a more gradual development over generations than other chains. Though this is only one example, it may suggest that, while transmission and interaction are general processes that can lead to systematic and communicatively efficient structures, they are not the only factors modulating how different forms evolve, and there may indeed be limits on the degree to which a structure can be systematized. For example, even in fully developed spatial modulation systems, signers make use of deictic points and iconic spatial mappings that may be highly variable between signers and across discourse contexts (Liddell 2003; Lillo-Martin and Meier 2011; Cormier et al. 2015; Fenlon et al. 2018; Schembri et al. 2018). Referential loci can occur at any location in the continuous space around the signer, and therefore provide an uncountably large set of possible locations to index (Liddell 2003). It proves difficult to systematically categorize a continuous space when there are few physical constraints on how that space can be used. Similarly, Padden (1986) argues that there are physical limitations on the number of roles that body-shift can contrast—two at the most. In sum, there are limitations within natural languages on how systematized such spatial mappings can become (Cormier et al. 2015; Fenlon et al. 2018), and these same restrictions may apply here. Future work could investigate specifically how different forms evolve under the same constraints to further understand the factors that are at play during communication and learning.

5 Conclusions

The study reported here has expanded on prior experimental research (e.g., Kirby, Cornish, and Smith 2008; Kirby et al. 2015; Silvey et al. 2015; Carr et al. 2016), investigating the effects of language use (i.e., interactive communication), and transmission (i.e., learning) by targeting the emergence of argument distinctions across a discourse. Here, we focused on the systematization of spatial modulation, a feature considered to be heavily modality-dependent, which has not been investigated experimentally until now. Participants in the experiment rely on a range of strategies to differentiate event arguments across discourse, some of which make use of space. The strategies participants use come to systematically refer to recurring agents through the use and transmission of these systems. The findings from this investigation support claims made on the basis of emerging sign languages such as ABSL, NSL, and ISL, namely that systematic spatial reference does not emerge fully formed in a language, but takes time to develop (Padden et al. 2010; Flaherty 2014; Montemurro et al. 2019). Furthermore, in the present study, it becomes systematized in ways that can facilitate comprehension by indicating critical contrasts between agents in target events. Our findings highlight the special role of iconicity in both encouraging and limiting the use of space in the manual modality, and offer insight into how highly iconic forms become integrated into a linguistic system.

Data availability

Both video and coded data from this experiment are publicly available. Video data can be accessed at: https://doi.org/10.7488/ds/2852. Recorded data from the experiments and coded data as well as all analysis scripts can be found at https://osf.io/hp5md/.

Supplementary material

Supplementary data is available at *Journal of Language Evolution* online.

Conflict of interest statement

No conflict of interest.

Funding

Y.M. was funded by a Carnegie PhD Scholarship from the Carnegie Trust for the Universities of Scotland (award number PHD060261). This project also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 681942), held by K.S., and an Economic and Social Research Council Research Grant held by S.K. and M.S. (award number ES/R011869/1).

Notes

1. Full description of the coding scheme and details of all agreement analyses can be found at https://osf.io/hp5md/. We also calculated Cohen's *d* where appropriate. As it is not an appropriate measure of agreement for all parameters, we report here percentage agreement.
2. Gestures for goals in target events will not be discussed here: goals were gestured with an explicit, identifiable gesture in only 29.5% of trials (excluding plain verb trials, where a goal is not encoded in the target event).
3. Coefficient estimates are similar in both models. The model including random slopes is reported in the supplementary materials.

Acknowledgements

We would like to thank Hanna Jarvinen and Danielle Naegeli, who second coded the data, and Pamela Perniss, who gave valuable feedback on an earlier draft of this manuscript. We also thank Parissa Motamedi, who provided illustrations for the manuscript.

References

Aronoff, M., Meir, I., and Sandler, W. (2005) 'The Paradox of Sign Language Morphology', *Language*, 81/2: 301–44

Bates, D. et al. (2015) 'Fitting Linear Mixed-Effects Models Using lme4', *Journal of Statistical Software*, 67/1: 1–48

Beckner, C., Pierrehumbert, J. B., and Hay, J. (2017) 'The Emergence of Linguistic Structure in an Online Iterated Learning Task', *Journal of Language Evolution*, 2/2 , 160–176.

Brennan, S. E., and Clark, H. H. (1996) 'Conceptual Pacts and Lexical Choice in Conversation', *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 22/6: 1482–93

Carr, J. W. et al. (2016) 'The Cultural Evolution of Structured Languages in an Open-Ended, Continuous World', *Cognitive Science*, 12371: 1–32

Chen Pichler, D. (2012) 'Acquisition' in Pfau, R., Steinbach, M., and Woll, B. (eds), *Sign Language: An International Handbook*, pp. 647–86. Berlin: De Gruyter Mouton.

Christensen, P., Fusaroli, R., and Tylén, K. (2016) 'Environmental Constraints Shaping Constituent Order in Emerging Communication Systems: Structural Iconicity, Interactive Alignment and Conventionalization', *Cognition*, 146: 67–80

Clark, H. H., and Wilkes-Gibbs, D. (1986) 'Referring as a Collaborative Process', *Cognition*, 22/1: 1–39

Coppola, M., and So, W. C. (2006) 'The Seeds of Spatial Grammar: Spatial Modulation and Coreference in Homesigning and Hearing Adults' in Bamman, D., Magnitskaia, T., and Zaller, C. (eds), *Proceedings of the Boston University Conference on Language Development*, Vol. 30, pp. 119–30. Boston, MA: Cascadilla Press.

Cormier, K., Fenlon, J., and Schembri, A. (2015) 'Indicating Verbs in British Sign Language Favour Motivated Use of Space', *Open Linguistics*, 1: 684–707

——, Smith, S., and Zwets, M. (2013) 'Framing Constructed Action in British Sign Language Narratives', *Journal of Pragmatics*, 55: 119–39

Dryer, M. S. (2013). 'Order of Subject, Object and Verb' in Dryer, M. S., and Haspelmath, M. (eds), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Fay, N., Arbib, M., and Garrod, S. (2013) 'How to Bootstrap a Human Communication System', *Cognitive Science*, 37/7: 1356–67

Fedzechkina, M., Jaeger, T. F., and Newport, E. L. (2012) 'Language Learners Restructure Their Input to Facilitate Efficient Communication', *Proceedings of the National Academy of Sciences of the United States of America*, 109/44: 17897–902

Fenlon, J., Schembri, A., and Cormier, K. (2018) 'Modification of Indicating Verbs in British Sign Language: A Corpus-Based Study', *Language*, 94/1, 84-119.

Ferrara, L., and Johnston, T. (2014) 'Elaborating Who's What: A Study of Constructed Action and Clause Structure in Auslan (Australian Sign Language)', *Australian Journal of Linguistics*, 34/2: 193–215

Flaherty, M. (2014). *The Emergence of Argument Structural Devices in Nicaraguan Sign Language. Doctoral Thesis*, University of Chicago, Chicago.

Garrod, S. et al. (2007) 'Foundations of Representation: Where Might Graphical Symbol Systems Come from?', *Cognitive Science*, 31/6: 961–87

Gibson, E. et al. (2019). 'How Efficiency Shapes Human Language', *Trends in Cognitive Sciences*, 23/5: 389–407.

—— et al. (2013) 'A Noisy-Channel account of Crosslinguistic Word-Order Variation', *Psychological Science*, 24/7: 1079–88

Goldin-Meadow, S. et al. (2008) 'The Natural Order of Events: How Speakers of Different Languages Represent Events Nonverbally', *Proceedings of the National Academy of Sciences of the United States of America*, 105/27: 9163–8

Hall, M. L., Mayberry, R. I., and Ferreira, V. S. (2013) 'Cognitive Constraints on Constituent Order: Evidence from Elicited Pantomime', *Cognition*, 129/1: 1–17

Johnston, T. A. (2019). 'Clause Constituents, Arguments and the Question of Grammatical Relations in Auslan (Australian Sign Language): A Corpus-Based Study', *Studies in Language*, 43/4: 941–96

Kirby, S. (2016). *Videobox*. University of Edinburgh: Edinburgh. Accessed May 20th 2021 at <http://edin.ac/2haREUz>

——, Cornish, H., and Smith, K. (2008) 'Cumulative Cultural Evolution in the Laboratory: An Experimental Approach to the Origins of Structure in Human Language', *Proceedings of the National Academy of Sciences of the United States of America*, 105/31: 10681–6

——, ——, and —— et al. (2015) 'Compression and Communication in the Cultural Evolution of Linguistic Structure Linguistic Structure', *Cognition*, 141: 87–102.

Kocab, A., Pyers, J., and Senghas, A. (2015) 'Referential Shift in Nicaraguan Sign Language: A Transition from Lexical to Spatial Devices', *Frontiers in Psychology*, 5:1540.

Leeson, L., and Saeed, J. (2012) Word Order. in Pfau, R., Steinbach, M. and Woll, B.(eds) *Sign Language: An International Handbook*, pp. 245-264. Berlin, Boston: De Gruyter Mouton.

Liddell, S. K. (2003) *Grammar, Gesture and Meaning in American Sign Language*. Cambridge: Cambridge University Press.

Lillo-Martin, D. (1999) 'Modality Effects and Modularity in Language Acquisition: The Acquisition of American Sign Language', in Bhatia, T., and Ritchie, W. (eds) *Handbook of Child Language Acquisition*, pp. 531–567. New York: Academic Press.

—— (2012). 'Utterance Reports and Constructed Action', in Pfau, R., Steinbach, M., and Woll, B. (eds) *Sign Language: An International Handbook*. Berlin, Boston: De Gruyter Mouton.

——, and Meier, R. (2011) 'On the Linguistic Status of 'agreement' in Sign Languages', *Theoretical Linguistics*, 37/3–4: 95–141

Mathur, G., and Rathmann, C. (2012). 'Verb Agreement', in Pfau, R., Steinbach, M., and Woll, B (eds) *Sign Language: An International Handbook*, pp. 136–57. Berlin: De Gruyter Mouton.

Meier, R. (1990). 'Person Deixis in American Sign Language', in Fischer, S., and Siple P.(eds) *Theoretical Issues in Sign Language Research*, pp. 175–90. Chicago: University of Chicago Press.

Meir, I. (2012) 'The Evolution of Verb Classes and Verb Agreement in Sign Languages', *Theoretical Linguistics*, 38/1–2: 145–52.

—— et al. (2014) 'The Effect of Being Human and the Basis of Grammatical Word Order: Insights from Novel Communication Systems and Young Sign Languages', *Cognition*, 158/April: 1–40.

—— et al. (2007) 'Body as Subject', *Journal of Linguistics*, 43/03: 531–63.

—— et al. (2013) 'Competing Iconicities in the Structure of Languages', *Cognitive Linguistics*, 24/2: 309–43.

Montemurro, K. et al. (2019). 'Grammaticalization of the Body and Space in Nicaraguan Sign Language', in Brown, M. M. and Dailey, B. (eds) *Proceedings of the 43rd Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Morgan, G., Barrière, I., and Woll, B. (2006) 'The Influence of Typology and Modality on the Acquisition of Verb Agreement Morphology in British Sign Language', First Language, 26/1, 19–43.

——, Herman, R., and —— (2002) 'The Development of Complex Verb Constructions in British Sign Language', *Journal of Child Language*, 29/3: 655–75

Motamedi, Y. et al. (2019) 'Evolving Artificial Sign Languages in the Lab: From Improvised Gesture to Systematic Sign', *Cognition*, 192: 103964

Napoli, D. J., and Sutton-Spence, R. (2014) 'Order of the Major Constituents in Sign Languages: Implications for All Language', *Frontiers in Psychology*, 5, 1–18

Nölle, J. et al. (2018) 'The Emergence of Systematicity: How Environmental and Communicative Factors Shape a Novel Communication System', *Cognition*, 181, 93–104

Nyst, V. A. (2007) *A Descriptive Analysis of Adamorobe Sign Language (ghana)*. Amsterdam: LOT.

Padden, C. (1986) 'Verbs and Role-Shifting in American Sign Language' in Padden, C. (ed.) *Proceedings of the Fourth National Symposium on Sign Language Research and Teaching*, pp. 44–57. Washington DC: The National Association for the Deaf.

—— (1990) 'The Relation between Space and Grammar in Asl Verb Morphology' in Lucas, C. (ed.) *Proceedings of the 2nd International Conference on Theoretical Issues in Sign Language Research*, pp. 118–32. Washington DC: Gallaudet University Press.

—— et al. (2010). The Grammar of Space in Two New Sign Languages. *Sign Languages: A Cambridge Language Survey*. Cambridge University Press: Cambridge, UK, pp. 570–92.

Peirce, J. W. (2007) 'Psychopy–Psychophysics Software in Python', *Journal of Neuroscience Methods*, 162/1–2: 8–13

Pfau, R., and Quer, J. (2010) 'Nonmanuals: Their Grammatical and Prosodic Roles' in Brentari, D. (ed.) *Sign Languages*, pp. 381–402. Cambridge: Cambridge University Press.

Quinto-Pozos, D. (2007) 'Can Constructed Action Be Considered Obligatory?', *Lingua*, 117/7: 1285–314.

R Core Team. (2008) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raviv, L., Meyer, A., and Lev-Ari, S. (2019) 'Compositional Structure Can Emerge without Generational Transmission', *Cognition*, 182: 151–64.

Roberts, G., Lewandowski, J., and Galantucci, B. (2015) 'How Communication Changes When we Cannot Mime the World: Experimental Evidence for the Effect of Iconicity on Combinatoriality', *Cognition*, 141: 52–66.

Sandler, W., and Lillo-Martin, D. (2006) *Sign Language and Linguistic Universals*. New York, NY: Cambridge University Press.

Schembri, A., Cormier, K., and Fenlon, J. (2018) 'Indicating Verbs as Typologically Unique Constructions: Reconsidering Verb 'Agreement' in Sign Languages', *Glossa: a Journal of General Linguistics*, 3/1: 89.

Schouwstra, M., and Swart, H. D. (2014) 'The Semantic Origins of Word Order', *Cognition*, 131/3: 431–6

Senghas, A. (2003) 'Intergenerational Influence and Ontogenetic Development in the Emergence of Spatial Grammar in Nicaraguan Sign Language', *Cognitive Development*, 18/4: 511–31

—— and Coppola, M. (2001) 'Children Creating Language: How Nicaraguan Sign Language Acquired a Spatial Grammar', *Psychological Science*, 12/4: 323–8.

Siewierska, A. (2013) 'Verbal Person Marking' in Dryer, M. S., and Haspelmath, M. (eds) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Silvey, C., Kirby, S., and Smith, K. (2015) 'Word Meanings Evolve to Selectively Preserve Distinctions on Salient Dimensions', *Cognitive Science*, 39/1: 212–26.

Smith, K. et al. (2016) 'Language Learning, Language Use, and the Evolution of Linguistic Variation', *Philosophical Transactions of the Royal Society B*, 1–16.

So, W. C. et al. (2005) 'The Seeds of Spatial Grammar in the Manual Modality', *Cognitive Science*, 29/6: 1029–43.

Sutton-Spence, R., and Woll, B. (1999). *The Linguistics of British Sign Language*. Cambridge: Cambridge University Press.

Taub, S. (2001). *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge: Cambridge University Press.

Theisen, C. A., Oberlander, J., and Kirby, S. (2010) 'Systematicity and Arbitrariness in Novel Communication Systems', *Interaction Studies*, 11/1: 14–32.

Theisen-White, C., Kirby, S., and Oberlander, J. (2011) 'Integrating the Horizontal and Vertical Cultural Transmission of Novel Communication Systems', in Carlson, L., Hoelscher, C., and Shipley, T. F. (eds) *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, pp. 956–61, Austin, TX: Cognitive Science Society.

Verhoef, T., ——, and Padden, C. (2011). 'Cultural Emergence of Combinatorial Structure in an Artificial Whistled Language', in Carlson, L., Hoelscher, C., and Shipley, T. F. (eds) *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, pp. 483–488. Austin, TX: Cognitive Science Society.

Vos, C. D. (2012). *Sign-Spatiality in Kata Kolok: How a Village Sign Language in Bali Inscribes Its Signing Space*. Unpublished doctoral dissertation, Radboud University Nijmegen.

Winters, J., Kirby, S., and Smith, K. (2014) 'Languages Adapt to Their Contextual Niche', *Language and Cognition*, 7/3: 415–449.